WO 00/53624
PCT/US00/06297

8/pts

CHEMICAL SYNTHESIS AND USE OF SOLUBLE MEMBRANE PROTEIN RECEPTOR DOMAINS

Introduction

Technical Field

The present invention relates to chemical synthesis and use of soluble extramembranous domains of membrane protein receptors.

Background

Neurotransmitters, peptide hormones, growth factors, and other molecules are ligands for cellular receptors that regulate signal transduction in and among cells, as well as the extracellular matrix. Most receptors are membrane proteins having extracellular, transmembrane and cytosolic domains. In the context of receiving and transduction of ligand-based extracellular signals, the general simplified function of these domains is as follows. The extracellular domain provides a ligand-binding site that receives information from outside the cell based on the presence or absence of the ligand. The transmembrane domain anchors the receptor protein within the plasma membrane and permits transduction of the information received by the extracellular domain to the cytosolic domain. The transmembrane domain of some receptors also may serve as a ligand-binding site. The cytosolic domain in turn transduces the signaling information received on the outside of the cell to the inside. Ligand-based information received from inside the cell via the cytosolic domain and/or the transmembrane domain may also contribute to the receptor-mediated signal transduction cascade.

There are many types of cellular receptors. Some receptors are at the center of signaling pathways that regulate changes in cellular events such as metabolism or gene expression in response to hormones and growth factors, while others affect cell adhesion and organization of the cytoskeleton. An example of a receptor family that effects cell adhesion and organization of the cytoskeleton is the family of integrin receptors. Integrin receptors are the major receptors responsible for the attachment of cells to the extracellular matrix. Most integrin receptors identified to date possess an extracellular domain that interacts with the extracellular matrix, an alpha-helix transmembrane domain, and a short cytoplasmic domain that lacks any intrinsic enzymatic activity.

Membrane protein receptors that regulate changes in cellular events such as metabolism or gene expression include enzyme-linked receptors. Enzyme-linked receptors are directly coupled to intracellular enzymes and include guanylyl cyclases, tyrosine kinases, tyrosine kinase-associated tyrosine phosphatases, and serine/threonine kinases receptors. The largest family of enzyme-linked receptors is the receptor protein-tyrosine kinases. This family includes receptors for epidermal growth factor (EGF), nerve growth factor (NGF), platelet-derived growth factor (PDGF), insulin and many other growth factors. Most enzyme-linked receptors have an N-terminal extracellular ligand binding domain, a single transmembrane alpha-helix domain, and a cytosolic C-terminal domain having tyrosine kinase activity. Binding of ligand such as growth factor to the extracellular domain activates the cytosolic kinase domain, which in turn propagates an intracellular signal. Binding of ligand to most enzyme-linked receptors induces dimerization of receptor monomer (e.g., EGF), whereas other receptors exist as dimers (e.g., insulin, PDGF and NGF receptors). For instance, ligand binding to receptors having a monomeric state crosslinks the monomers and induces dimerization.

Cytokine receptors and non-receptor protein kinases represent another family of enzyme-linked membrane protein receptors. This family includes receptors for cytokines such as interleukin-2 and erythropoietin, as well as for some polypeptide hormones such as growth hormone. These receptors have an N-terminal extracellular ligand binding domain, a single transmembrane alpha-helix domain, and a cytosolic C-terminal domain. They differ from protein-tyrosine kinase receptors in that the C-terminal domain does not by itself posses kinase activity. Instead, these receptors transmit ligand-binding information through intracellular protein kinases associated with the C-terminal cytosolic domain.

The largest family of cell surface receptors transmits signals to intracellular targets via the intermediary action of guanine nucleotide-binding proteins called G-proteins. More than a thousand such G-protein coupled receptors (GPCRs) have been identified to date. GPCRs are characterized by seven transmembrane domains that terminate in an extracellular N-terminal domain and an intracellular or cytoplasmic C-terminal domain. Thus, GPCRs also have been referred to as "7TM" receptors. The GPCRs can be classified into three major subfamilies related to rhodopsin (type A), calcitonin (type B), and metabolic receptors (type C).

Another family of membrane protein receptors is the ion channel receptors. The ion channel receptors include ligand-gated and voltage-gated ion channels. Ligand-gated ion channels are pentamers of homologous subunits. The acetylcholine receptor is an example of a ligand-gated ion channel. Voltage-gated ion channels are homotetramers with subunits having six transmembrane helices. Potassium and sodium ion channels are examples of voltage-gated ion channels. Extracellular and/or intracellular domains that bind ligand are likely to be important in regulating many ion channels.

Cellular membrane protein receptors represent extremely important drug targets. For example, ion channels are therapeutic targets for major human diseases such as cardiac arrhythmias, stroke, hypertension, heart failure, asthma, diabetes, cystic fibrosis, epilepsy, migraine, and depression. Enzyme-linked receptors are implicated in multiple disorders including diabetes, cancer, and blood and nervous system disorders. Cytokine receptors are therapeutic targets for immune system disorders such as AIDS and arthritis. GPCRs are recognized as the largest groups of receptors targeted by commercially available drugs. For instance, GPCR type B receptors are important drug targets for mediation of metabolic disease, nervous system disorders, cancer and other diseases. Family members include receptors for glucagon, glucagon-like peptide, VIP (vasoactive intestinal polypeptide), GIP (gastro-intestinal peptide), GHRH (growth-hormone releasing hormone), secretin, PACAP (pituitary adenyl cyclase activating polypeptide), PTH (parathyroid hormone), calcitonin and CRF (corticotropin-releasing factor). In addition, the type B GPCR family includes different subtypes and several orphan receptors.

Drug targets typically include those related to ligand binding, since drugs can be employed that modulate natural ligand interaction with its receptor. As mentioned above, most ligand binding sites of receptors reside in the extramembranous portions of receptors, such as the extracellular and cytosolic domains. Unfortunately, very little is known about the structure- and quantitative structure-activity relationship (SAR/QSAR) for most membrane protein receptors, including their extramembranous domains. This lack of information has hampered development of new drugs and the understanding of these molecules in general. By way of example, even though recombinant forms of GPCRs have been made, including isolated domains, detailed structure/function information for the extramembranous domains of these receptors

remains elusive. For instance, the N-terminal extramembranous domain of type B GPCRs appears to be highly crosslinked by disulfide bridges formed by six well-conserved cysteine residues. Replacement of any one of the six cysteine residues, reduction of the receptor with beta-mercaptoethanol, and deletion of the N-terminal domain strongly reduces the binding affinity for the respective hormone ligand in several family members (Wilmen, et al., *FEBS Letters* (1996) 398:43-47; DeAlmeida, et al., *Molecular Endocrinology* (1998) 12:750-765). Also, site-directed mutagenesis experiments on the VIP receptor have identified several conserved residues (Asp67, Trp72, Pro86, Gly108 and Trp110) besides the cysteines that appear to be crucial for receptor activity in some members of the family (Couvineau, et al., *Biochem. Biophys. Res. Comm.* (1995) 206:246-252; and Wilmen, et al., *Peptides* (1997) 18:301-305). Even though theoretical modeling, and *in vitro* and *in vivo* assays have been utilized, the disulfide-pairing pattern of the crucial disulfide bonds of type B GPCRs still has yet to be established. This can be attributed in large part to the difficulty in producing and purifying sufficient quantities and true homogenous preparations of materials needed for such studies (Willshaw, et al., *Biochemical Society Transactions* (1998) 26:S288; and Chow, et al., *Recept. Signal Transduct.* (1997) 7:143-150).

To date, production of extracellular and cytosolic domains of membrane protein receptors has all but been limited to recombinant expression of the domains joined to at least one transmembrane anchoring region (See, e.g., Hsuesh, et al., WO 97/39131). Otherwise very little product can be made, much less isolated in useful amounts. Nevertheless, it is unlikely that a recombinantly produced domain of a receptor can be purified to an extent that it represents a true homogenous material, or that is free of cellular contaminants, which is a problem with all recombinant expression systems no matter how stringent and redundant the purification conditions might be. Another frustration with recombinantly produced receptors is that they cannot be, for all practical purposes, site-specifically labeled at any position within the molecule, particularly cytosolic and extracellular portions of a receptor in isolation from the transmembrane spanning domain(s). Labeling, for instance, has been restricted to conjugation of labels to only a few amino acids in the full length receptor following expression (See, e.g., Gether, et al., *J. Biol. Chem.* (1995) 270:28268-28275; and Kim, et al., *Biochemistry* (1998) 37(13):4680-4686), incorporation of radioactive or spin labels throughout the receptor or by use of *in vitro* suppression

mutagenesis in *Xenopus* oocytes of full length receptor (see, e.g., Turcatti, et al., *J. Bio. Chem.* (1996) 271:19991-19998). Nor does recombinant expression by itself provide access to ultra pure and large amounts of ultra homogenous and functional extramembranous receptor domain material. The present invention addresses these and other problems.

Relevant Literature

Wilken, et al. (*Curr. Opin. Biotech.* (1998) 9(4):412-426) review chemical protein synthesis. Turcatti, et al. (*J. Bio. Chem.* (1996) 271:19991-19998) disclose *in vitro* suppression mutagenesis in *Xenopus* oocytes to introduce fluorescence-labeled amino acids into the seven transmembrane neurokinin-2 receptor and its incorporation and activity in oocyte membranes. Gether, et al., (*J. Biol. Chem.* (1995) 270:28268-28275) disclose fluorescent labeling of cysteine residues in the transmembrane domain of the beta-2-andreogenic receptor. Hsuesh, et al. (WO 97/39131) disclose recombinant expression of a transmembrane anchor coupled through a protease cleavage site to the N-terminal portion of a G-protein coupled receptor. Various references disclose recombinant expression of N-terminal domains of GPCRs (Willshaw, et al., *Biochemical Society Transactions* (1998) 26:S288, Wilmen, et al., *FEBS Letters* (1996) 398:43-47; Chow, et al. (*Receptor Signal Transduction* (1997) 7:143-150; and Cao, et al., *Biochem Biophys Res. Commun.* (1995) 212(2):673-680 disclose recombinant expression of the secretin/VIP N-terminal domain). Bozon, et al. (*J. Mol. Endo.* (1995) 14:227) and Bobovnikova, et al. (*Endocrinology* (1995) 138:588) report on recombinant expression of leutonizing hormone (LH) and thyroid stimulating hormone (TSH) receptor domains, respectively. U.S. Patent Nos. 5,726,290 and 5,837,486 discloses soluble analogs of integrins and assays. U.S. Patent Nos. 5,783,402 and 5,462,856 disclose cell-based GPCR-linked assays for ligands.

Summary Of The Invention

The invention relates to chemical synthesis of extramembranous receptor domains of membrane protein receptors, and compositions and methods that employ them. The extramembranous receptor domains include the soluble ligand-binding extracellular and cytosolic domains of membrane protein receptors. The extramembranous receptor domains of the invention are produced by ligating under chemoselective chemical ligation conditions first and second peptides of an

extramembranous receptor domain of a membrane protein receptor, where the peptides have unprotected chemoselective reactive groups capable of forming a covalent bond therein between. The ligation product is exposed to a folding buffer having a chaotropic reagent and an organic solvent that approximates the water-lipid interface of a cell membrane. Exposure to the folding buffer is followed by isolation from the buffer of ligation product that binds to a ligand of the membrane protein receptor. The ligand-binding portion of the ligation product produced by this method represents folded extramembranous receptor domain.

Also provided is a composition comprising a synthetic extramembranous receptor domain of a membrane protein receptor having a chemically synthesized segment that includes an unnatural amino acid at a pre-selected residue position, where the extramembranous receptor domain is free of a membrane spanning transmembrane domain and is capable of binding to a ligand of the membrane protein receptor. Compositions having a totally synthetic and ultra homogenous extramembranous receptor domain of a membrane protein receptor free of cellular contaminants also are provided.

The invention further includes a method of detecting binding of a ligand to a soluble extramembranous receptor domain of a membrane protein receptor. This aspect of the invention involves contacting the monomer of a soluble extramembranous receptor domain of a membrane protein receptor with a ligand of the membrane protein receptor, where the soluble extramembranous receptor domain is free of a membrane spanning transmembrane domain and includes an unnatural amino acid at a pre-selected residue position. The contacting is followed by assaying the soluble extramembranous receptor domain for ligand-induced association of domain monomers, such as dimerization.

The invention also includes a method of assaying a soluble extramembranous receptor domain monomer for ligand-induced association of domain monomers. This method includes contacting a soluble extramembranous receptor domain of a membrane protein receptor with a ligand of the membrane protein, where the soluble extramembranous receptor domain is free of a membrane spanning transmembrane domain and includes an unnatural amino acid at a pre-selected residue position. The contacting is followed by assaying the soluble extramembranous receptor domain for ligand-induced association of domain monomers.

Also provided is a method of detecting binding of a ligand to an extramembranous receptor domain of a membrane protein receptor. This method involves contacting a soluble extramembranous receptor domain of a membrane protein receptor with a ligand for the membrane protein receptor, where the soluble extramembranous receptor domain is free of a membrane spanning transmembrane domain and comprises an unnatural amino acid having a detectable moiety. Detection of ligand binding is then performed by assaying for a change in a property of the detectable moiety, such as fluorescence when the detectable label is a fluorophore.

The methods and compositions of the invention provide unprecedented access to non-limiting amounts of ultra pure and ultra homogenous soluble extramembranous receptor domains of membrane protein receptors, including extracellular and cytosolic domains having one or more unnatural amino acids and/or unnatural reactive functional groups at pre-selected positions. This facilitates for the first time production and true site-specific labeling of soluble membrane receptor domains free of impurities, transmembrane spanning regions, and other characteristic of domains made solely by recombinant synthesis. The invention also provides synthetic access to structure/function information previously unattainable by other approaches, as well as discovery of novel information about extramembranous domains of receptors for exploitation in drug discovery, disease treatment and diagnostics. The methods and compositions of the invention are particularly useful for high throughput screening of compounds that are ligands for the receptors corresponding to the extramembranous receptor domains.

Definitions

Amino Acid: Include the 20 genetically coded amino acids, rare or unusual amino acids that are found in nature, and any of the non-naturally occurring and modified amino acids. Sometimes referred to as amino acid residues when in the context of a peptide, polypeptide or protein.

Chemoselective Chemical Ligation: Chemically selective reaction involving covalent ligation of (1) a first unprotected amino acid, peptide or polypeptide with (2) a second amino acid, peptide or polypeptide. Any chemoselective reaction chemistry that can be applied to ligation of unprotected peptide segments.

Extramembranous Receptor Domain: A domain of a membrane protein receptor that is external to a lipid membrane bilayer of a cell. Includes extracellular

and cytosolic domains of a membrane protein receptor, or soluble portions of these domains that are capable of binding to a ligand of the membrane protein receptor, such as N-terminal and C-terminal extramembranous domains. Excludes membrane spanning transmembrane domain that anchors an intact membrane protein receptor to the lipid membrane bilayer of a cell. Can include one or more amino acid residues of a transmembrane domain, provided the residues do not span the membrane or form an insoluble complex incapable of binding to a ligand.

Ligand: A chemical entity that interacts with its target membrane protein receptor of a cell.

Membrane Protein Receptor: A receptor of a cell having at least one peptide segment capable of being embedded and anchoring the receptor in the lipid bilayer of a cell membrane. Examples include, by way of illustration and not limitation, enzyme-linked receptors including receptor protein-tyrosine kinases such as receptors for EGF, NGF, PDGF, insulin and many other growth factors; and cytokine receptors and non-receptor protein kinases such as receptors for cytokines such as interleukin-2 and erythropoietin, as well as for some polypeptide hormones such as growth hormone; G-protein coupled receptors including those related to rhodopsin (type A), calcitonin (type B), and metabolic receptors (type C), more particularly including type B receptors such as receptors for glucagon, glucagon-like peptide, VIP, GIP, GHRH, secretin, PACAP, PTH and CRF; and ion channel receptors including ligand-gated channels such as the acetylcholine receptor and voltage-gated ion channels such as the potassium and sodium ion channels.

Peptide: A polymer of at least two monomers, wherein the monomers are amino acids, sometimes referred to as amino acid residues, which are joined together via an amide bond. May have either a completely native amide backbone or an unnatural backbone or a mixture thereof. Can be prepared by known synthetic methods, including solution synthesis, stepwise solid phase synthesis, segment condensation, and convergent condensation. Can be synthesized ribosomally in cell or in a cell free system, or generated by proteolysis of larger polypeptide segments. Can be synthesized by a combination of chemical and ribosomal methods.

Polypeptide: A polymer comprising three or more monomers, wherein the monomers are amino acids, sometimes referred to as amino acid residues, which are joined together via an amide bond. Also referred to as a peptide or protein. Can

comprise native amide bonds or any of the known unnatural peptide backbones or a mixture thereof. Range in size from 3 to 1000 amino acid residues, preferably from 3-100 amino acid residues, more preferably from 10-60 amino acid residues, and most preferably from 20-50 amino acid residues. Segments or all of the polypeptide can be prepared by known synthetic methods, including solution synthesis, stepwise solid phase synthesis, segment condensation, and convergent condensation. Segments or all of the polypeptide also can be prepared ribosomally in a cell or in a cell-free translation system, or generated by proteolysis of larger polypeptide segments. Can be synthesized by a combination of chemical and ribosomal methods.

Protein Domain: A contiguous stretch of amino acid residues within a protein sequence related to a functional property of the molecule.

Soluble Membrane Protein Receptor Domain: A domain of a membrane protein receptor that is soluble under aqueous physiologic conditions. Examples include the extracellular and intracellular domains of a receptor that reside in an aqueous environment external or internal to a lipid membrane bilayer of a cell, respectively.

Brief Description Of The Drawings

Fig. 1 is a schematic showing the extramembranous receptor domains of a G-coupled protein receptor, in the context of the intact membrane protein receptor.

Fig. 2 is a schematic showing the extramembranous receptor domains of a G-coupled protein receptor incorporating various unnatural amino acids, in the context of the intact membrane protein receptor. "X" represents a label, "Y" represents an unnatural backbone, and "Z" represents a chemical handle.

Fig. 3 is a schematic showing an N-terminal extracellular receptor domain of a Type B G-coupled protein receptor.

Fig. 4 is a schematic showing the chemical ligation design for the N-terminal extracellular receptor domain of a Type B Glucagon-like peptide 1 G-coupled protein receptor (GLP-1R), using Segment 1 (SEQ ID NO:1), Segment 2 (SEQ ID NO:2) and Segment 3 (SEQ ID NO:3).

Fig. 5 shows folding of the chemically synthesized GLP-1R N-terminal domain as monitored using mass spectroscopy and high performance liquid chromatography (HPLC).

Figs. 6A and 6B show chymotryptic digest of the chemically synthesized GLP-1R N-terminal domain and mapping of disulfide bonds.

Fig. 7 shows the disulfide bond map of the chemically synthesized GLP-1R N-terminal domain (SEQ ID NO:4).

Figs. 8A and 8B show binding of rhodamine-labeled GLP ligand to the GLP-1R N-terminal domain and fluorescence anisotropy measurement of the binding event.

Description Of Specific Embodiments

The invention relates to the chemical synthesis of extracellular receptor domains of membrane protein receptors. and compositions and methods that employ them. The extramembranous receptor domains include the soluble ligand-binding extracellular and intracellular domains of membrane protein receptors. The extramembranous receptor domains of the invention are produced by first selecting the domain targeted for synthesis. Amino acid sequence information of the domain is then utilized to design peptide or polypeptide segments for chemical ligation. Peptide segments of a selected extramembranous receptor domain are then constructed so as to have unprotected chemoselective reactive groups capable of forming a covalent bond therein between when contacted under conditions amenable to the chosen method of chemical ligation. The peptide segments are then covalently joined by chemoselective chemical ligation. The ligation product formed by chemical ligation of peptide segments is then exposed to a folding buffer to generate a folded ligation product. The folding buffer contains at least one chaotropic reagent and an organic solvent that mimics the water-lipid interface environment of a cell membrane. Exposure to the folding buffer is followed by isolation from the buffer of ligation product that binds to a ligand of the membrane protein receptor. such as a natural ligand of the receptor. The ligand-binding portion of the ligation product produced by this method represents the folded extramembranous receptor domain.

Selection of an extramembranous receptor domain can be accomplished by identifying and retrieving amino acid sequence information for the receptor or domain to be synthesized. such as from a private or public database. Examples of public accessible databases useful for this purpose include. for example. GenBank (Benson. et al.. *Nucleic Acids Res.* (1998) 26(1):1-7); USA National Center for Biotechnology Information. National Library of Medicine (National Institutes of Health. Bethesda.

MD. USA). TIGR Database (The Institute for Genomic Research. Rockville. MD. USA). Protein Data Bank (Brookhaven National Laboratory. USA). and the ExPASy and Swiss-Protein database (Swiss Institute of Bioinformatics. Geneva. Switzerland). Alternatively. the amino acid sequence information can be obtained *de novo* using standard biochemical and/or molecular biology techniques. such as cloning and sequencing following protocols well known in the art. When one or more target extramembranous domains of a chosen receptor sequence have not been defined, then various rules of thumb. screening and/or modeling techniques known in the art can be employed to characterize putative domains. Examples include sequence homology comparisons. and use of algorithms and protein modeling tools known in the art that are suitable for this purpose. For instance. mutagenesis. thermodynamic, computational. modeling and/or any technique that reveals functional and/or structural information regarding a target polypeptide of interest can be used for this process. These techniques include immunological and chromatographic analyses, fluorescence resonance energy transfer (FRET), circular dichroism (CD). nuclear magnetic resonance (NRM). electron and x-ray crystallography. electron microscopy, Raman laser spectroscopy and the like. which are commonly exploited for designing and characterizing membrane polypeptide systems. (See. e.g., Newman. R., *Methods Mol. Biol.* (1996) 56:365-387; Muller, et al.. *J. Struct. Biol.* (1997) 119(2):149-157; Fleming, et al., *J. Mol. Biol.* (1997) 272: 266-27: Haltia. et al., *Biochemistry* (1994) 33(32): 9731-9740.5; Swords. et al.. *Biochem. J.* (1993) 289(1): 215-219; Wallin, et al., *Protein Sci.* (1997) 6(4):808-815; Goormaghtigh. et al., *Subcell Biochem.* (1994) 23:405-450). Muller. et al.. *Biophys. J.* (1996) 70(4):1796-1802; Sami. et al.. *Biochim Biophys Acta.* (1992) 1105(1):148-154. Wang. et al., *J. Mol. Biol.* (1994) 237(1):1-4; Watts, et al., *Mol. Membr. Biol.* (1995) 12(3):233-246; Bloom. M.. *Biophys. J.* (1995) 69(5):1631-1632; and Gutierrez-Merino. et al.. *Biochem Soc. Trans.* (1994) 22(3):784-788).

Since most receptors typically are identified by their characteristic transmembrane spanning domains. these regions can be used as a reference to identify non-transmembrane segments that are likely to exist external to the cell membrane. In particular. structural and functional information can be obtained using standard techniques including homology comparisons to other membrane protein receptors having similar amino acid sequences and domains. preferably other receptors for

which at least some structural and functional information is known. For an exemplary list of membrane protein receptors of known three-dimensional structure, see Preusch, et al., *Nature Struct. Biol.* (1998) 5:12-14.

Modeling programs also may be employed to identify putative transmembrane segments, and thus putative extramembranous domains. For example, the programs "TmPred" and TopPredII" can be used to make predictions of membrane-spanning regions and their orientation, which is based on the statistical analysis of a database of transmembrane proteins present in the SwissProt database. (von Heijne, *J. Mol. Biol.* (1992) 225:487-494; Hoppe-Seyler, *Biol. Chem.* (1993) 347:166; and Claros, et al., *Comput. Appl. Biosci.* (1994) 10(6):685-686). Other programs can be used and include: "DAS" (Cserzo, et al., *Prot. Eng.* (1997) 10(6):673-676); "PHDhtm" (Rost, et al., *Protein Science* (1995) 4:521-533); and "SOSUI" (Mitaku Laboratory, Department of Biotechnology, Tokyo University of Agriculture and Technology).

A target receptor also may be modeled in three dimensions to identify putative extramembranous receptor domains therein. First, a sequence alignment between the polypeptide to be modeled and a polypeptide of known structure is established. Second, a backbone structure is generated based on this alignment. This is normally the backbone of the most homologous structure, but a hybrid backbone also may be used. Third, sidechains are then placed in the model. Various techniques like Monte Carlo procedures, tree searching algorithms etc., can be used to model rotomer sidechains having multiple possible conformations. If the polypeptide to be modeled has insertions or deletions with respect to the known structure, loops are re-modeled, or modeled *ab initio*. Database searches for loops with similar anchoring points in the structure are often used to build these loops, but energy based *ab initio* modeling techniques also can be employed. Energy minimizations, sometimes combined with molecular dynamics, are then normally used for optimization of the final structure. The quality of the model is then assessed, including visual inspection, to verify that the structural aspects of the model are not contradicting what is known about the functional aspects of the molecule.

The three-dimensional models are preferably generated using a computer program that is suitable for modeling membrane polypeptides. (Vriend, "Molecular Modeling of GPCRs" in 7TM (1995) vol. 5). Examples of computer programs suitable for this purpose include: "WhatIf" (Vriend, *J. Mol. Graph.* (1990) 8:52-56;

available from EMBL, Meyerhofstrasse 1, 69117 Heidelberg , Germany) and "Swiss-Model" (Peitsch, et al., (1996) "Molecular modeling of G-protein coupled receptors" in G Protein-coupled Receptors. New opportunities for commercial development, 6:6.29-6.37, N Mulford and LM Savage Eds., IBC Biomedical Library Series; Peitsch, et al., *Receptors and Channels* (1996) 4:161-164; Peitsch, et al., "Large-scale comparative protein modeling" in Proteome research: new frontiers in functional genomics," pp. 177-186, Wilkins MR, Williams KL, Appel RO, Hochstrasser DF, Eds., Springer, 1997).

Amino acid sequence information of the selected extramembranous domain is then utilized to design peptide or polypeptide segments for chemical ligation, where at least one peptide employed for ligation is chemically synthesized, i.e., via ribosomal free synthesis. In developing the ligation strategy, peptide segments are designed to provide unprotected reactive groups that selectively react to yield a covalent bond at the ligation site, also referred to as a chemoselective ligation site. Thus the peptides are designed in accordance with a selected ligation chemistry, or more than one individual ligation chemistry, provided the segments targeted for ligation in any given step of synthesis provide compatible chemoselective ligation component pairings that form the desired covalent bond upon chemoselective chemical ligation, which avoids unwanted side reactions.

In particular, peptide or polypeptide segments and their chemoselective reactive groups are designed based on the ligation chemistry selected for covalently stitching the segments together in their desired orientation. Any number of ligation chemistries may be employed in accordance with the methods of the invention. These chemistries include, but are not limited to, native chemical ligation (Dawson, et al., *Science* (1994) 266:776-779; Kent, et al., WO 96/34878), extended general chemical ligation (Kent, et al., WO 98/28434), oxime-forming chemical ligation (Rose, et al., *J. Amer. Chem. Soc.* (1994) 116:30-33), thioester forming ligation (Schnölzer, et al., *Science* (1992) 256:221-225), thioether forming ligation (Englebretsen, et al., *Tet. Letts.* (1995) 36(48):8871-8874), hydrazone forming ligation (Gaertner, et al., *Bioconj. Chem.* (1994) 5(4):333-338), thiazolidine forming ligation and oxazolidine forming ligation (Zhang, et al., *Proc. Natl. Acad. Sci.* (1998) 95(16):9184-9189; Tam, et al., WO 95/00846).

By way of example, for native chemical ligation, the ligation component segments for ligation comprise a compatible native chemical ligation component pairing in which one of the components provides a cysteine having an unprotected amino group and the other component provides an amino acid having an unprotected α-thioester group. These groups are capable of chemically reacting to yield a native peptide bond at the ligation site. For oxime-forming chemical ligation, the peptide segments comprise a compatible oxime-forming chemical ligation component pairing in which one of the components provides an unprotected amino acid having an aldehyde or ketone moiety and the other component provides an unprotected amino acid having an amino-oxy moiety. These groups are capable of chemically reacting to yield a ligation product having an oxime moiety at the ligation site. For thioester-forming chemical ligation, the ligation segments comprise a compatible thioester-forming chemical ligation component pairing in which one of the components provides an unprotected amino acid having a haloacetyl moiety and the other component provides an unprotected amino acid having an α-thiocarboxylate moiety. These groups are capable of chemically reacting to yield a ligation product having a thioester moiety at the ligation site. For thioether-forming chemical ligation, the ligation components comprise a compatible thioether-forming chemical ligation component pairing in which one of the components provides an unprotected amino acid having a haloacetyl moiety and the other component provides an unprotected amino acid having an alkyl thiol moiety. These groups are capable of chemically reacting to yield a ligation product having a thioether moiety at the ligation site. For hydrazone-forming chemical ligation, the ligation components comprise a compatible hydrazone-forming chemical ligation component pairing in which one of the components provides an unprotected amino acid having an aldehyde or ketone moiety and the other component provides an unprotected amino acid having an hydrazine moiety. These groups are capable of chemically reacting to yield a ligation product having a hydrazone moiety at the ligation site. For thiazolidine-forming chemical ligation, the ligation components comprise a compatible thiazolidine-forming chemical ligation component pairing in which one of the components provides an unprotected amino acid having a 1-amino, 2-thiol moiety and the other component provides an unprotected amino acid having an aldehyde or a ketone moiety. These groups are capable of chemically reacting to yield a ligation product having a

thiazolidine moiety at the ligation site. For oxazolidine-forming chemical ligation, the ligation components comprise a compatible oxazolidine-forming chemical ligation component pairing in which one of the components provides an unprotected amino acid having a 1-amino, 2-hydroxyl moiety and the other component provides an unprotected amino acid presenting an aldehyde or a ketone moiety. These groups are capable of chemically reacting to yield a ligation product having an oxazolidine moiety at the ligation site.

Other design considerations include uniqueness of the ligation site, solubility of the ligation components, and specificity and completeness of the ligation reaction. In particular, ligation component pairings are preferably designed to maximize selectivity of the ligation reaction. This includes design of linker or capping sequences that may be employed to generate chemoselective reactive groups, or used to assist in the solubility of the ligation components. The ligation components and complementary pairings thereof also can be selected by modeling them in two or three-dimensions to simulate the ligated product and/or the pre-ligation reaction components. Modeling programs as described above can be used for this purpose.

When designing a smaller extramembranous receptor domain, all peptides can be synthesized chemically and employed for total chemical synthesis and ligation. This includes, for example, domains ranging in size up to about 200 to 250 amino acids. These totally synthetic domains also may be ligated together to form even larger totally synthetic domains. Since chemical synthesis is utilized, the chemically synthesized peptides or polypeptides can be linear, cyclic or branched, and often composed of, but not limited to, the 20 genetically encoded L-amino acids. Chemical synthetic approaches also permit incorporation of novel or unusual chemical moieties including D-amino acids, other unnatural amino acids, oxime, hydrazone, ether, thiazolidine, oxazolidine, ester or alkyl backbone bonds in place of the normal amide bond, N- or C-alkyl substituents, side chain modifications, and constraints such as disulfide bridges and side chain amide or ester linkages. See, for example, Wilken, et al., (*Curr. Opin. Biotech.* (1998) 9(4):412-426), which reviews various chemistries for chemical synthesis of peptides and polypeptides.

For example, native chemical ligation and synthesis of polypeptides having a native peptide backbone structure is disclosed in Kent, et al., WO 96/34878. See also Dawson, et al. (*Science* (1994) 266:77-779) and Tam, et al. (*Proc. Natl. Acad. Sci.*

USA (1995) 92:12485-12489). Unnatural peptide backbones also can be made by known methods (See, e.g., Schnolzer, et al., *Science* (1992) 256:221-225; Rose, et al., *J. Am Chem. Soc.* (1994) 116:30-34; Liu, et al., *Proc. Natl. Acad. Sci. USA* (1994) 91:6584-6588; Englebretsen, et al., *Tet. Letts.* (1995) 36(48):8871-8874; Gaertner, et al., *Bioconj. Chem.* (1994) 5(4):333-338; Zhang, et al., *Proc. Natl. Acad. Sci.* (1998) 95(16):9184-9189; and Tam, et al., WO 95/00846). Extended general chemical ligation and synthesis also may be employed as disclosed in Kent, et al., WO 98/28434.

Additionally, rapid methods of synthesizing assembled polypeptides via chemical ligation of three or more unprotected peptide segments using a solid support, where none of the reactive functionalities on the peptide segments need to be temporarily masked by a protecting group, and with improved yields and facilitated handling of intermediate products is described in Canne, et al., WO 98/56807. Briefly, this method involves solid phase sequential chemical ligation of peptide segments in an N-terminus to C-terminus direction, with the first solid phase-bound unprotected peptide segment bearing a C-terminal α-thioester that reacts with another unprotected peptide segment containing an N-terminal cysteine and a C-terminal thioacid. The techniques also permits solid-phase native chemical ligation in the C- to N-terminus direction. Large polypeptides can also be synthesized by chemical ligation of peptide segments in aqueous solution on a solid support without need for protecting groups on the peptide segments. A variety of peptide synthesizers are commercially available for batchwise and continuous flow operations as well as for the synthesis of multiple peptides within the same run and are readily automated.

For larger extramembranous receptor domains, it may be desirable to chemically synthesize one or more smaller peptide or polypeptide segments that incorporate an unnatural or modified amino acid having a selected reactive group (R1) at a chosen termini, and utilize one or more recombinantly produced polypeptide segments that include a terminal amino acid having a selected reactive group (R2), where R1 and R2 represent compatible chemoselective reactive groups. An example is utilization of native chemical ligation in which the recombinant segment is provided with a terminal cysteine residue (R2) that is capable of chemoselective ligation to a thioester provided by the selected reactive group (R1) of the chemically synthesized peptide segment.

Some commonly used host cell systems for cloning, expression and recovery of membrane protein receptor polypeptides include *E. coli*, *Xenopus* oocytes, baculovirus, vaccinia, and yeast, as well as many higher eukaryotes including transgenic cells in culture and in whole animals and plants. (See, e.g., G.W. Gould, "Membrane Protein Expression Systems: A User's Guide," Portland Press, 1994, Rocky S. Tuan, ed.; and "Recombinant Gene Expression Protocols," Humana Press, 1996). For example, yeast expression systems are well known and can be used to express and recover a target membrane protein receptor polypeptide of interest following standard protocols. (See, e.g., Nekrasova, et al, *Eur. J. Biochem.* (1996) 238:28-37; Gene Expression Technology Methods in Enzymology 185 (1991); Molecular Biology and Genetic Engineering of Yeasts CRC Press, Inc. (1992); Herescovics, et al., *FASEB* (1993) 7:540-550; Larriba, G. *Yeast* (1993) 9:441-463; Buckholz, R.G., *Curr. Opinion Biotech.* (1993) 4:538-542; Asenjo, et al., "An Expert System for Selection and Synthesis of Protein Purification Processes Frontiers in Bioprocessing II" pp. 358-379, American Chemical Society, (1992); Mackett, M, "Expression of Membrane Proteins in Yeast Membrane Protein Expression Systems: A Users Guide" pp. 177-218, Portland Press (1995)).

Cleavage sites also may be suitably positioned into the segment utilized for ligation, so that cleavage yields the desired terminal group for ligation. Some commonly encountered protease cleavage sites are: Thrombin (KeyValProArg/GlySer); Factor Xa Protease (IleGluGlyArg); Enterokinase (AspAspAspAspLys); rTEV (GluAsnLeuTyrPheGln/Gly), which is a recombinant endopeptidase from the Tobacco Etch Virus; and 3C Human rhino virus Protease (LeuGluValLeuPhe Gln/GlyPro) (Pharmacia Biotech).

Various chemical cleavage sites are also known and include, but are not limited to, the intein protein-splicing elements (Dalgaard, et al., *Nucleic Acids Res.* (1997) 25(6):4626-4638) and cyanogen bromide cleavage sites. Inteins can be constructed which fail to splice, but instead cleave the peptide bond at either splice junction (Xu, et al., *EMBO J.* (1996) 15(19):5146-5153; and Chong, et al., *J. Biol. Chem.* (1996) 271:22159-22168). For example, the intein sequence derived from the *Saccharomyces cerevisiae* VMA1 gene can be modified such that it undergoes a self-cleavage reaction at its N-terminus at low temperatures in the presence of thiols such

as 1,4-dithiothreitol (DTT), 2-mercaptoethanol or cysteine (Chong, et al., *Gene* (1997) 192:271-281).

Cyanogen bromide (CnBr) cleaves at internal methionine (Met) residues of a polypeptide sequence. Cleavage with CnBr yields two or more fragments, with the fragments containing C-terminal residues internal to the original polypeptide sequence having an activated alpha-carboxyl functionality, e.g. cyanogen bromide cleavage at an internal Met residue to give a fragment with a C-terminal homoserine lactone. For some polypeptides, the fragments will re-associate under folding conditions to yield a folded polypeptide-like structure that promotes reaction between the segments to give a reasonable yield (often 40-60%) of the full-length polypeptide chain (now containing homoserine residues where there were Met residues subjected to cyanogen bromide cleavage) (Woods, et al., *J. Biol. Chem.*, (1996), 271:32008-32015).

In general, a cleavage site for generating a ligation site amenable to a desired chemical ligation chemistry usually is selected to be unique, i.e., it occurs only once in the target polypeptide. However, when more than one cleavage site is present in a target polypeptide that is recognized and cleaved by the same cleavage reagent, if desired one or more of such sites can be permanently or temporarily blocked from access to the cleavage reagent and/or removed during synthesis. Cleavage sites can be removed during synthesis of a given peptide segment by replacing, inserting or deleting one or more residues of the cleavage reagent recognition sequence, and/or incorporating one or more unnatural amino acids that achieve the same result (See Figs. 1 and 2). A cleavage site also may be blocked by agents that bind to the peptide, including ligands that bind the peptide and remove accessibility to all or part of the cleavage site. However a cleavage site is blocked or removed, one of ordinary skill in the art will recognize that the method is selected such that upon cleavage the peptide or polypeptide is capable of chemoselective chemical ligation to a target ligation component of interest.

A ligation component also can be selected to contain moieties that facilitate and/or ease purification and/or detection. For example, purification handles or tags that bind to an affinity matrix can be used for this purpose (See Figs. 1 and 2). Many such moieties are known and can be introduced via post-synthesis chemical modification and/or during synthesis. (See, e.g., Protein Purification Protocols,

(1996), Doonan, ed., Humana Press Inc.; Schriemer, et al., *Anal Chem.* (1998) 70(8):1569-1575; Evangelista, et al., *J. Chromatogr. B. Biomed. Sci. Appl.* (1997) 699(1-2):383-401; Kaufmann, *J. Chromatogr. B. Biomed. Sci. Appl.* (1997) 699(1-2):347-369; Nilsson, et al, *Protein Expr. Purif.* (1997) 11(1):1-16; Lanfermeijer, et al., *Protein Expr. Purif.* (1998) 12(1): 29-37). For example, one or more unnatural amino acids having a chemical moiety that imparts a particular property that can be exploited for purification can be incorporated during synthesis. Purification sequences also can be incorporated by recombinant DNA techniques. In some instances, it may be desirable to include a chemical or protease cleavage site to remove the tag, depending on the tag and the intended end use. An unnatural amino acid or chemically modified amino acid also may be employed to ease detection, such as incorporation of a chromophore, hapten or biotinylated moiety detectable by fluorescence spectroscopy, immunoassays, and/or MALDI mass spectrometry.

Thus, one or more of the peptides or polypeptides utilized for ligation may be (1) totally synthetic, i.e., produced *in toto* by ribosomal free chemical synthesis; (2) semi-synthetic, i.e., produced at least partially using ribosomal synthesis such as via recombinant DNA techniques; or (3) natural, i.e., produced *in toto* by ribosomal synthesis. The extramembranous receptor domains of the invention can thus be totally synthetic or semi-synthetic, and may include one or more unnatural amino acids incorporated at pre-selected residue positions, as illustrated in Figs. 1 and 2.

Once the extramembranous receptor domain is designed and the ligation components prepared, the segments are ligated under the appropriate chemical ligation conditions corresponding to the chosen ligation chemistry(s) imparted by the design. As can be appreciated, reaction conditions for a given ligation chemistry are selected to maintain the desired interaction of the ligation components. For example, pH and temperature, and solubilizing reagents can be varied to optimize ligation. Addition or exclusion of reagents that solubilize the ligation components to different extents may further be used to control the specificity and rate of the desired ligation reaction. Reaction conditions are readily determined by assaying for the desired chemoselective reaction product compared to one or more internal and/or external controls.

Homogeneity and the structural identity of the desired ligation products can be confirmed by any number of means including immunoassays, fluorescence

spectroscopy, gel electrophoresis. HPLC using either reverse phase or ion exchange columns. amino acid analysis. mass spectrometry. crystallography. NMR and the like. Positions of amino acid modifications. insertions and/or deletions. if present. can be identified by sequencing with either chemical methods (Edman chemistry) or tandem mass spectrometry.

For folding. the ligation product is dissolved in an aqueous solution containing a solubilizing amount of a chaotropic reagent at a pH compatible with the ligation product in question. Chaotropic reagents suitable for this purpose include. for example. urea and guanidinium chloride. The concentration of the chaotropic reagent and pH of the solution can be adjusted for optimal solubilization of the ligation product. but within a range that does not damage the protein. When cysteines are present in the ligation product. a disulfide reducing agent such as dithiothreitol may be used. The solution containing the dissolved ligation product is then diluted by admixing with folding buffer. The folding buffer includes a buffering reagent, a chaotropic reagent and an organic solvent that are combined in amounts that mimic the water-lipid interface of a cell membrane. Buffering reagents are well known and include salts. such as Tris and Mops. The organic solvent utilized in the folding buffer is chosen to have a chemical moiety that hydrogen bonds with water, and another chemical group providing an aliphatic moiety. Preferred organic solvents are water soluble. Examples of water soluble organic solvents include monohydroxy alcohols such as methyl, ethyl. n-propyl, isopropyl, tert-butyl, and allyl alcohols. Diol and triol alcohols such as ethylene glycol. propylene glycol. trimethyl glycol and glycerol also may be utilized. Preferred water soluble organic solvent for use in the methods of the invention are methanol and glycerol, with the most preferred being methanol. It will be appreciated that organic solvents and other folding buffer components that denature proteins are avoided. or are present in non-denaturing amounts. It also will be appreciated that one or more chaotropes. organic solvents and the like can be employed for folding. Other additives may be included such as detergents. lipids and the like. This includes chaperone proteins. Moreover, the folding buffer may be utilized for folding of the ligation product in a perfusion device, for example. as with the oxidative refolding chromatography approach described in Altamirano. et al. (*Nature Biotechnology* (1999) 17:187-191). that employs an immobilized chaperone system. The ligation buffer also may include additives such

as reduced and/or oxidized glutathione, for example, when disulfide bond forming cysteines are present. It also will be appreciated that the actual components and ratios thereof employed in the folding buffer of the method of the invention can be determined for a given ligation product, and adjusted as necessary. The ligation product is exposed to the folding buffer for a period of time sufficient to permit folding, which can be monitored by any number of standard techniques described above and known in the art. A preferred monitoring method is liquid chromatography, e.g. HPLC, which also permits isolation of folding products concurrent with monitoring. Folding products are separated by any number of non-denaturing chromatographic techniques, and then assayed for binding to a ligand of the membrane protein from which the extramembranous receptor domain was derived. Assays known in the art and/or those described herein can be used for this purpose. Folding product that binds to the ligand is then categorized as a folded extramembranous receptor domain. The folded ligation product can then be utilized immediately or stored, in unfolded or folded form for later use.

In another embodiment of the invention, compositions are provided that are produced according to the method of the invention. One composition of the invention includes a synthetic extramembranous receptor domain of a membrane protein receptor having a chemically synthesized segment that includes an unnatural amino acid at a pre-selected residue position, where the extramembranous receptor domain is free of a membrane spanning transmembrane domain and is capable of binding to a ligand of the membrane protein receptor. Compositions of the invention also include a totally synthetic and ultra homogenous extramembranous receptor domain of a membrane protein receptor free of cellular contaminants. The extramembranous receptor domain of the compositions of the invention may comprise one or more synthetic segments that include genetically encoded L-amino acids, linear, cyclic or branched amino acids, D-amino acids, other unnatural amino acids, as well as oxime, hydrazone, ether, thiazolidine, oxazolidine, ester or alkyl backbone bonds in place of the normal amide bond, N- or C-alkyl substituents, side chain modifications, and constraints such as disulfide bridges and side chain amide or ester linkages.

Preferred unnatural amino acids are those having a detectable label. In this embodiment, chemical synthesis is utilized to incorporate at least one detectable label in a pre-ligation component. In this way the resulting ligation product can be

designed to contain one or more detectable labels at pre-specified positions of choice. Isotopic labels detectable by NMR are of particular interest. Also of particular interest is the incorporation of one or more unnatural amino acids comprising a detectable label at one or more specific sites in a target ligation component of interest. By unnatural amino acid is intended any of the non-genetically encoded L-amino acids and D-amino acids that are modified to contain a detectable label, such as photoactive groups, as well as chromophores including fluorophores and other dyes, or a hapten such as biotin. Unnatural amino acids comprising a chromophore and chemical synthesis techniques used to incorporate them into a peptide or polypeptide sequence are well known, and can be used for this purpose. For example, it may be convenient to conjugate a fluorophore to the N-terminus of a resin-bound peptide before removal of other protecting groups and release of the labeled peptide from the resin. Fluorescein, eosin, Oregon Green, Rhodamine Green, Rhodol Green, tetramethylrhodamine, Rhodamine Red, Texas Red, coumarin and NBD fluorophores, the dabcyl chromophore and biotin are all reasonably stable to hydrogen fluoride (HF), as well as to most other acids, and thus suitable for incorporation via solid phase synthesis. (Peled, et al., *Biochemistry* (1994) 33:7211; Ben-Efraim, et al., *Biochemistry* (1994) 33:6966). Other than the coumarins, these fluorophores also are stable to reagents used for deprotection of peptides synthesized using FMOC chemistry (Strahilevitz, et al., *Biochemistry* (1994) 33:10951). The *t*-BOC and α-FMOC derivatives of ε-dabcyl-L-lysine also can be used to incorporate the dabcyl chromophore at selected sites in a polypeptide sequence. The dabcyl chromophore has broad visible absorption and can used as a quenching group. The dabcyl group also can be incorporated at the N-terminus by using dabcyl succinimidyl ester (Maggiora, et al., *J. Med. Chem.* (1992) 35:3727). EDANS is a common fluorophore for pairing with the dabcyl quencher in FRET experiments. This fluorophore is conveniently introduced during automated synthesis of peptides by using 5-((2-(t-BOC)-γ-glutamylaminoethyl) amino) naphthalene-1-sulfonic acid (Maggiora, et al., *supra*). An α-(t-BOC)-ε-dansyl-L-lysine can be used for incorporation of the dansyl fluorophore into polypeptides during chemical synthesis (Gauthier, et al., *Arch Biochem Biophys* (1993) 306:304). As with EDANS fluorescence of this fluorophore overlaps the absorption of dabcyl. Site-specific biotinylation of peptides can be achieved using the t-BOC-protected derivative of biocytin (Geahlen, et al., *Anal.*

Biochem. (1992) 202:68), or other well known biotinylation derivatives such as NHS-biotin and the like. Racemic benzophenone phenylalanine analog also can be incorporated into peptides following its t-BOC or FMOC protection (Jiang, et al., *Intl. J. Peptide Prot. Res.* (1995) 45:106). Resolution of the diastereomers can be accomplished during HPLC purification of the products: the unprotected benzophenone also can be resolved by standard techniques in the art. Keto-bearing amino acids for oxime coupling, aza/hydroxy tryptophan, biotyl-lysine and D-amino acids are among other examples of unnatural amino acids that can be utilized. It will be recognized that other protected amino acids for automated peptide synthesis can be prepared by custom synthesis following standard techniques in the art.

It will be appreciated that other detectable labels can be incorporated into a ligation component post-chemical ligation, although less preferred. This can be done by chemical modification using a reactive substance that forms a covalent linkage once having bound to a reactive group of the target molecule. For example, a peptide or polypeptide ligation component can include several reactive groups, or groups modified for reactivity, such as thiol, aldehyde, amino groups, suitable for coupling the detectable label by chemical modification (Lundblad, et al., in "Chemical Reagents for Protein Modification", CRC Press, Boca Raton, FL, (1984)). Site-directed mutagenesis and/or chemical synthesis also can be used to introduce and/or delete such groups from a desired position. Any number of detectable labels including biotinylation probes of a biotin-avidin or streptavidin system, antibodies, antibody fragments, carbohydrate binding domains, chromophores including fluorophores and other dyes, lectin, nucleic acid hybridization probes, drugs, toxins and the like, can be coupled in this manner. For instance, a low molecular weight hapten, such a fluorophore, digoxigenin, dinitrophenyl (DNP) or biotin, can be chemically attached to the membrane polypeptide or ligation label component by employing haptenylation and biotinylation reagents. The haptenylated polypeptide then can be directly detected using fluorescence spectroscopy, mass spectrometry and the like, or indirectly using a labeled reagent that selectively binds to the hapten as a secondary detection reagent. Commonly used secondary detection reagents include antibodies, antibody fragments, avidins and streptavidins labeled with a fluorescent dye or other detectable marker.

Depending on the reactive group, chemical modification can be reversible or irreversible. A common reactive group targeted in peptides and polypeptides are thiol groups, which can be chemically modified by haloacetyl and maleimide labeling reagents that lead to irreversible modifications and thus produce more stable products. For instance, reactions of sulfhydryl groups with α-haloketones, amides, and acids in the physiological pH range (pH 6.5-8.0) are well known and allow for the specific modification of cysteines in peptides and polypeptides (Hermason, et al., in "Bioconjugate Techniques", Academic Press, San Diego, CA, pp. 98-100, (1996)). Covalent linkage of a detectable label also can be triggered by a change in conditions, for example, in photoaffinity labeling as a result of illumination by light of an appropriate wavelength. For photoaffinity labeling, the label, which is often fluorescent or radioactive, contains a group that becomes chemically reactive when illuminated (usually with ultraviolet light) and forms a covalent linkage with an appropriate group on the molecule to be labeled. An important class of photoreactive groups suitable for this purpose is the aryl azides, which form short-lived but highly reactive nitrenes when illuminated. Flash photolysis of photoactivatable or "caged" amino acids also can be used for labeling peptides that are biologically inactive until they are photolyzed with UV light. Different caging reagents can be used to modify the amino acids, such derivatives of o-nitrobenzylic compounds, and detected following standard techniques in the art. (Kao, et al., "Optical Microscopy: Emerging Methods and Applications," B. Herman, J.J. Lemasters, eds., pp. 27-85 (1993)). The nitrobenzyl group can be synthetically incorporated into the biologically active molecule via an ether, thioether, ester (including phosphate ester), amine or similar linkage to a heteroatom (usually O, S or N). Caged fluorophores can be used for photoactivation of fluorescence (PAF) experiments, which are analogous to fluorescence recovery after photobleaching (FRAP). Those caged on the ε-amino group of lysine, the phenol of tyrosine, the γ-carboxylic acid of glutamic acid or the thiol of cysteine can be used for the specific incorporation of caged amino acids in the sequence. Alanine, glycine, leucine, isoleucine, methionine, phenylalanine, tryptophan and valine that are caged on the α-amine also can be used to prepare peptides that are caged on the N-terminus or caged intermediates that can be selectively photolyzed to yield the active amino acid either in a polymer or in solution. (Patchornik, et al., *J. Am. Chem. Soc.* (1970) 92:6333). Spin labeling

techniques of introducing a grouping with an unpaired electron to act as an electron spin resonance (ESR) reporter species may also be used, such as a nitroxide compound (-N-O) in which the nitrogen forms part of a sterically hindered ring (Oh, et al., *supra*).

Selection of a detectable label system generally depends on the assay and its intended use. In particular, the chemical ligation methods and compositions of the invention can be employed in a screening or detection assay of the invention. These include diagnostic assays, screening new compounds for drug development, and other structural and functional assays that employ binding of a ligand to a extramembranous receptor domain produced by the method of the invention. The ligands may be derived from naturally occurring ligands or derived from synthetic sources, such as combinatorial libraries. Screening and detection methods of particular interest involve detection of ligand binding by fluorescence spectroscopy.

In one embodiment, a soluble extramembranous receptor domain of a membrane protein receptor produced by the method of the invention is utilized to detect binding of a ligand thereto. This aspect of the invention involves contacting monomers of a soluble extramembranous receptor domain of a membrane protein receptor with a ligand of the membrane protein receptor. The soluble extramembranous receptor domain used in this method is free of a membrane spanning transmembrane domain and includes an unnatural amino acid at a pre-selected residue position, such as an unnatural amino acid comprising a detectable label. The contacting is followed by assaying the soluble extramembranous receptor domain for ligand-induced association of domain monomers. For example, association of domain monomers, such as dimerization, can be detected by monitoring a change in the property of the detectable label.

In another embodiment, a method of assaying a soluble extramembranous receptor domain monomer for ligand-induced association of domain monomers is provided. This method includes contacting a soluble extramembranous receptor domain of a membrane protein receptor with a ligand of the membrane protein receptor. In this method, as in the above method, the soluble extramembranous receptor domain is free of a membrane spanning transmembrane domain and includes an unnatural amino acid at a pre-selected residue position. The contacting is followed

by assaying the soluble extramembranous receptor domain for ligand-induced association of domain monomers.

In yet another embodiment, a method of detecting binding of a ligand to an extramembranous receptor domain of a membrane protein receptor is provided. This method involves contacting a soluble extramembranous receptor domain of a membrane protein receptor with a ligand for the membrane protein receptor, where the soluble extramembranous receptor domain is free of a membrane spanning transmembrane domain and comprises an unnatural amino acid having a detectable moiety. Detection of ligand binding is then performed by assaying for a change in a property of the detectable moiety, such as fluorescence when the detectable label is a fluorophore.

Of particular interest are methods and compositions employing a totally synthetic N-terminal extramembranous domain of a GPCR, such as a type B GPCR exemplified in the Examples. For instance, very little is known about the structure of type B GPCRs and few homogenous and truly high-throughput assays exist. The following gives a list of possible applications of synthetic N-terminal receptor domain in drug-discovery.

N-terminal receptor domains with FRET probes ligand displacement assays

In this assay format the N-terminal receptor domain and its ligand are labeled with a fluorescent donor and acceptor, respectively. Binding of the ligand to receptor will result in energy transfer between the ligand and the receptor. Small molecules that disrupt this interaction can be identified due to their interference with the energy transfer. Time-resolved luminescent probes, such as the lanthanide chelator complexes are ideal for this purpose, since they allow to reject background signals due to light-scattering and are compatible with current homogenous high-throughput screening equipment. Depending on the binding stoichiometry, other FRET assay formats can be envisioned. An intriguing possibility is that the binding of the hormone to the receptor results in receptor dimerization (or oligomerization). This means that one can envision receptor agonists that act by inducing dimerization (oligomerization) of the receptor.

Dimerization (oligomerization) assay

For dimerization (oligomerization) assays employing the N-terminal domains of GPCRs, in this assay format, one labels a portion of the receptor domains with a

fluorescent acceptor and the other half with a fluorescent donor. In the absence of a ligand inducing dimerization (oligomerization), no FRET is observed. However, the presence of such a ligand is indicated by a rise in a FRET signal. Time-resolved luminescent probes, such as lanthanide chelator complexes are ideal for this purpose, since they allow to reject background signals due to light-scattering and are compatible with current homogenous high-throughput screening equipment. Donor-Donor dimerized (oligomerized) pairs will not contribute to acceptor emission. Acceptor-Acceptor pairs can be gated away using an appropriate time delay. This leaves an unambiguous contribution from the Donor-Acceptor pair only.

Receptor domain labeled with isotopic NMR probes

Structure/Activity Relationship by Nuclear Magnetic Resonance (SAR by NMR) is a novel approach to drug screening that requires isotopic labeling of the drug target protein to identify interactions of small molecule drug precursors with a drug target. This approach detects changes in the chemical shift of an amino acid located in the binding site of a drug target that is induced by binding of a potential agonist or antagonist. Current SAR by NMR approaches require 2-dimensional NMR techniques on homogeneously isotopically labeled proteins, placing a severe constraint on the throughput of molecules amenable for screening. Chemical synthesis of proteins uniquely allows for the site-specific incorporation of isotopic labels into large quantities of protein, potentially requiring only 1-dimensional NMR techniques for SAR by NMR. This will provide significant time-savings per sample, propelling SAR by NMR into the realm of true high-throughput screening.

N-terminal receptor binding domains for phage display screening

Phage display is a very sensitive technique that allows for the amplification and identification of peptides that bind to an immobilized drug target. Chemical synthesis techniques are uniquely suited to generate large quantities of ion channels with site-specific attachment sites, e.g. via biotin labeling. Attachment of such a labeled domain to a solid support can then be used to select for phages that display a peptide exhibiting binding affinity to the N-terminal receptor domain. This will allow for the rapid identification of peptides that bind to a specific N-terminal receptor domain and that can be used as lead compounds for drug discovery. This approach could also be used to identify ligands for orphan receptors.

N-terminal receptor domains on a support matrix

As described for phage display attachment, an N-terminal receptor domain having a chemical handle such as a biotin label, can be used to attach the protein to a support matrix such as chip or a polymer. Such a device could be used to identify small peptides binding to an N-terminal receptor domain. In this approach, one synthesizes a library of small peptides. A solution of these peptides is incubated with the matrix. Unbound peptides are then washed off. Peptides binding to the receptor domain can then be easily analyzed by MALDI analysis. The same approach could also be used to identify ligands for orphan receptors.

Structural information for structure-based drug design

In addition, chemically synthesized receptor domains could be used to gain structural information that is crucial for structure based drug design. Such information includes NMR and crystallographic data on the free and ligand-bound state as well as structural data of the receptor domain complexed with a novel agonist or antagonist. Crystallographic studies will be aided by synthesis of N-terminal receptor domains with heavy isotope labels (such as selenomethionine and iodotyrosine).

Therapeutic Uses

The receptor domains of the present invention also find use as therapeutic agents, including use as agonists or antagonists of the corresponding naturally occurring receptor ligands, and as vaccines.

As an example, the GPCR type B receptor domains of the present invention can be utilized in the mediation of metabolic disease, nervous system disorders, cancer and other disease indications associated with the GPCR type B receptors. Indeed, there is an emerging sense that soluble forms of receptor domains represent an important new class of protein therapeutics for the treatment of human diseases (See, e.g., Heaney, et al., *J. Leukocyte Biology* (1998) 64(2):135-46). Some specific examples include recombinantly expressed soluble proteins containing a soluble IL-6 (Interleukin 6) receptor domain has been shown to act as agonists of IL-6 and normal IL-6 receptor activity (Mackiewicz, et al., *FEBS* (1992) 30:257). Accordingly, it is envisioned that synthetic IL-6 receptor domains produced according to the methods of the invention can be utilized as agonists on IL-6 receptor signaling. As another example, a proinflammatory cytokine, tumor necrosis factor alpha (TNFα) is involved in mediation of acute and chronic inflammation, and recombinant antibody-like

proteins comprising a TNFα receptor binding domain have found use as ligand traps for TNFα (Edwards CK 3rd., *Annals of the Rheumatic Diseases*. 1999 Nov. 58 Suppl 1:I73-81; Solorzano. et al., *J. Appl. Phys.* (1998) 84(4):1119-1130). Thus, synthetic TNFα receptor domains produced according to the methods of the invention can be used as antagonists to treat chronic inflammatory disease associated with the TNFα inflammation pathway. Furthermore, the synthetic receptor domains of the present invention can be utilized in a clinical application to generate neutralizing antibodies for blocking a membrane receptor protein involved in disease or for use as a vaccine. For instance, inhibition of the IL-2 receptor via a neutralizing antibody produces an effect having immunotherapeutic value (Rosenberg, S.A., *Immunology Today* (1988) 9:58-59). Also, the extracellular domain of the minor, virus-coded M2 protein is nearly invariant in all influenza A strains. Administration of a fusion proteins of the M2 domain to the hepatitis B virus core (HBc) protein provided 90-100% protection to mice against otherwise deadly viral infection (Neirynck, et al., *Nature Medicine* (1999) 5(10):1157-1163). One may thus utilize the methods and synthetic receptor domains of the invention as a broadband influenza vaccine constructs made up of the extracellular domain of the influenza A M2 protein as well as the homologous protein domains for influenza B and C joined in a multivalent fashion through a linker, or in a template assisted synthetic protein (TASP) construct design.

Given the absolute precision and power of chemical synthesis in constructing ultra pure and homogenous receptor domains compounds according to the present invention, these compounds thus find use in clinical applications that cannot be addressed using recombinant DNA techniques alone.

The following Examples are intended to illustrate various aspects of the invention and are not intended to limit the scope of the invention.

Examples

Example 1
Peptide Synthesis

The following peptide segments (See Fig. 3) for chemical synthesis of the GLP-1 receptor N-terminal domain (GLP-1R NTD) were synthesized using a custom-modified Applied Biosystems 430A peptide synthesizer following established protocols (Schnölzer, et al., *Int. J. Peptide Protein Res.* (1992) 40:180-193).

Segment 1 (SEQ ID NO:1):

AGPRPQGATVSLWETVQKWREYRRQCQRSLTEDPPPATDLF

Segment 2 (SEQ ID NO:2):

CNRTFDEYACWPDGEPGSFVNVSCPWYLPWASSVPQGHVYRF

Segment 3 (SEQ ID NO:3):

CTAEGLWLQKDNSSLPWRDLSECEESKRGERSSPEEQLLFL

A putative signaling domain consisting of 20 amino acid residues (Göke, et al., *FEBS Letters* (1996) 398:43-47) was conveniently excluded in the synthesis design. Peptide segments were purified by preparative gradient reversed phase HPLC on a Rainin dual-pump high-pressure mixing system with 214 nm UV detection using a Vydac C-4 preparative or semi-preparative column (10 mm particle size, 2.2 cm x 25 cm, and 1 cm x 25 cm, respectively) and analytical reversed phase HPLC was performed on a Vydac C-18 analytical column (5 mm particle size, 0.46 cm x 15 cm), using a Hewlett Packard Model 1100 quaternary pump high-pressure mixing system with 214 nm and 280 nm UV detection. Electrospray mass spectra (ESMS) of the peptide products were obtained using a PE-Sciex API-1 quadrupole ion-spray mass spectrometer. Peptide masses were calculated from all the observed protonation states and peptide mass spectra were reconstructed using the MACSPEC software (PE-Sciex, Thornhill, ON, Canada). Theoretical masses were calculated using the MACPROMASS software (Terri Lee, City of Hope). GLP-1R NTD segments 1 (SEQ ID NO:1) and 2 (SEQ ID NO:2) (amino acid residues 21-61 and 62-103, respectively) were synthesized on a thioester generating resin by the *in situ* neutralization protocol for Boc (tert-butoxycarbonyl) chemistry stepwise SPPS (Schnölzer, et al., *supra*), using established side-chain protection strategies. The N-terminal cysteine of segment 2 was protected with an ACM (acetamidomethyl) group to prevent cyclization. GLP-1R NTD segment 3 (SEQ ID NO:3) (amino acid residues 104-144) was synthesized analogously on a -OCH_2-PAM resin (Schnölzer, et al., *supra*). The peptides were deprotected and simultaneously cleaved from the resin support using HF/p-cresol according to standard Boc-chemistry procedures (Schnölzer, et al., *supra*). All three GLP-1R NTD segments were purified by preparative reversed-phase HPLC with a linear gradient of 25-45% Buffer B (100% acetonitrile containing 0.1% TFA) versus 0.1% aqueous TFA in 45 minutes. Fractions containing pure peptide were identified using ESMS, pooled and lyophilized for subsequent ligation.

The purified peptides were characterized by electrospray MS: Segment 1 thioester peptide (SEQ ID NO:1) obs. MW: 4962 + 1 D, calc. MW : 4,961.3 D (average isotope composition); Segment 2 thioester peptide with 2 protecting groups (1 His(DNP) and 1 Cys(ACM) groups) (SEQ ID NO:2): obs. MW 5,294 + 1 kD , calc. MW 5,294.4 D (average isotope composition); Segment 3 carboxylate peptide (SEQ ID NO:3) obs. MW: 4769 + 1 D, calc. MW : 4,749.3 D (average isotope composition).

Example 2
Chemical Protein Synthesis

Equimolar amounts of the purified unprotected GLP-1R NTD peptide (amino acid residues 104-144, Segment 3, SEQ ID NO:3) was added to a solution of the purified unprotected thioester peptide GLP-1R NTD (amino acid residues 62-103)alpha-COSR (Segment 2, SEQ ID NO:2) (2 mM) in 0.1M sodium phosphate/6M guanidinium chloride, pH 7.5 and 1% thiophenol. The ligation mixture was stirred overnight at room temperature and the reaction was monitored by reversed-phase HPLC and ESMS. The reaction mixture was subsequently treated with an equal volume of a solution of 40% beta-mercaptoethanol in 6M guanidinium chloride, 100 mM phosphate, pH 7.5) for 20 minutes to remove any residual His(DNP) protecting groups. Reactants and products were separated by preparative reversed-phase HPLC with a linear gradient of 25-45% Buffer B versus 0.1% aqueous TFA in 45 minutes. Fractions containing GLP-1R NTD (amino acid residues 62-144, Segments 2 and 3, SEQ ID NOS:2 and 3) were identified by ESMS (obs. MW 9,690 + 1 kD , calc. MW 9690.7 D (average isotope composition)), pooled and lyophilized. Subsequently, the purified GLP-1R NTD (62-144) was dissolved in 0.5 M acetic acid containing 2M urea and a 1.5 molar excess (relative to the total cysteine concentration) of $Hg(acetate)_2$. After 30 minutes, the solution was made 20% in beta-mercaptoethanol to scavenge mercury ions. Subsequently, the solution was desalted by preparative reversed-phase HPLC with a step gradient of 10-45% Buffer B versus 0.1% aqueous TFA and the resulting lyophilized GLP-1R NTD (amino acid residues 62-144, Segments 2 and 3, SEQ ID NOS: 2 and 3).

Equimolar amounts of the purified unprotected GLP-1R NTD (amino acid residues 62-144, Segments 2 and 3, SEQ ID NOS: 2 and 3) was added to a solution of the purified unprotected thioester peptide GLP-1R NTD(21-61)alpha-COSR (Segment 1, SEQ ID NO:1) (2 mM) in 0.1 M sodium phosphate/6M guanidinium

chloride, pH 7.5 and 1% thiophenol. Ligation and workup proceeded as described above to generate the following ligation product:

Synthetic GLP-1R N-terminal domain (amino acid residues 21-144, SEQ ID NO:4):

AGPRPQGATVSLWETVQKWREYRRQCQRSLTEDPPPATDLF
CNRTFDEYACWPDGEPGSFVNVSCPWYLPWASSVPQGHVYR
FCTAEGLWLQKDNSSLPWRDLSECEESKRGERSSPEEQLLFL

Reactants and products were separated by preparative reversed-phase HPLC with a linear gradient of 25-45% Buffer B versus 0.1% aqueous TFA in 45 minutes. Fractions containing full-length GLP-1R NTD (amino acid residues 21-144, SEQ ID NO:4) were identified by ESMS (obs. MW 14,375 + 1 kD , calc. MW 14,375.0 D (average isotope composition)), pooled and lyophilized.

Example 3
Protein Folding

The purified full-length GLP-1R NTD (amino acid residues 21-144, SEQ ID NO:4) was dissolved at 4 mg/ml in freshly degassed 6M guanidinium/HCl, pH 4 (100 mM sodium acetate) under an argon atmosphere. A 1 molar equivalent of DTT (dithiothreitol) was added and the solution was stirred for 30 min. Subsequently, the solution was diluted to a peptide concentration of 0.2 mg/ml with freshly degassed 2M guanidinium/HCl, pH 8.6 (200 mM Tris) containing 20% methanol, and an 8 molar equivalent of reduced glutathione and a 1 molar equivalent of oxidized glutathione (equivalents to cysteine concentration in the peptide) was added (Wetlaufer, et al., *Biochemistry* (1970) 9(25):5015). The solution was stirred under argon overnight. The progress of folding was monitored by analytical reversed-phase HPLC with a linear gradient of 25-45% Buffer B versus 0.1% aqueous TFA for 30 minutes until no change in the shape of the HPLC-trace was detected and most of the protein peaks had collapsed under one main peak, suggesting homogenous folding. The formation of 3 disulfide bridges during folding was identified by ESMS.

The folded full-length GLP-1R NTD (amino acid residues 21-144, SEQ ID NO:4) was purified by preparative reversed-phase HPLC with a linear gradient of 25-45% Buffer B versus 0.1% aqueous TFA in 45 minutes. Fractions containing folded full-length GLP-1R NTD (amino acid residues 21-144, SEQ ID NO:4) were identified by ESMS (obs. MW 14,369 + 1 kD , calc. MW 14,369.0 D (average isotope composition)), pooled and lyophilized.

Example 4
Analysis of Folded Product

Fig. 7 presents the sequence of the GLP-1R NTD (SEQ ID NO:4). Strictly conserved residues in type-B GPCR's are bolded, as well as the ligation sites and all cysteines, including the ligation sites at Cys62 and Cys104. Ligation of the 3 segments at these sites provides efficient access to multi-mg quantities of full-length GLP-1R NTD (21-144) peptide. Fig. 5 shows analytical reversed-phase HPLC traces monitoring the folding reaction of GLP-1R NTD. The top trace presents an HPLC-trace of full-length GLP-1R NTD (21-144) peptide dissolved in 6M guanidinium chloride, pH 4. The sharp peak at 21.1 minutes suggests high purity of the synthetic unfolded material. After an hour, multiple peaks in the HPLC-trace indicate a wide range of folding intermediates (data not shown). After overnight folding, most of these folding intermediates have disappeared and the corresponding HPLC-peaks have collapsed into the main peak at 18.9 minutes. A broader peak at earlier retention time is due to glutathion adduct formation. The formation of 3 disulfide bridges in the folded protein was confirmed by the observation of a mass loss of 6 D relative to the unfolded protein (See insets; unfolded full-length GLP-1R NTD (amino acid residues 21-144, SEQ ID NO:4) obs. MW 14,375 + 1 D , calc. MW 14,175.0 D (average isotope composition)), folded full-length GLP-1R NTD (amino acid residues 21-144, SEQ ID NO:4) obs. MW 14,369 + 1 D, calc. MW 14,369 D (average isotope composition)). Folded protein was separated from unfolded protein by reversed phase HPLC. Re-dissolving the lyophilized, folded protein gave solutions that showed GLP-1 binding activity.

Example 5
Proteolytic Digest of Folded Product & Disulfide Mapping

For proteolytic digest, 50 µg folded GLP-1R NTD (21-144) was dissolved in 100 µl 125 mM Tris-HCl, pH 7.5 containing 2 M urea and 10 mM $CaCl_2$. 4.5 µg CLCK treated chymotrypsin (49 u/g, Worthington Biochemicals) was added. The solution was stirred for 1 hour under argon and acidified with 100 µl 200 mM aqueous acetic acid. The peptide mixture was separated by analytical reversed-phase HPLC with a linear gradient of 5-45% Buffer B. Individual peptide fragments were identified by electrospray mass spectroscopy. Electrospray mass spectra (ESMS) of the digestion peptide products were obtained using a PE-Sciex API-1 quadrupole ion-spray mass spectrometer. Peptide masses were calculated from all the observed

protonation states and peptide mass spectra were reconstructed using the MACSPEC software (PE-Sciex, Thornhill, ON, Canada).

Figs. 6A and 6B present the results of the chymotryptic digest of the digested folded protein prior and after treatment of the digest mixture with TCEP (tris carboxyethylphosphine) and the difference spectrum between the chromatograms. In the difference chromatogram, one clearly observes positive peaks for the disulfide bonded peptide fragments containing disulfide bridges between Cys94 and Cys126 (amino acid residues 70-80 of SEQ ID NO: 4 and amino acid residues 81-87 of SEQ ID NO: 4), and Cys71 and Cys85 (amino acid residues 104-110 of SEQ ID NO: 4 and amino acid residues 121-142 of SEQ ID NO: 4). Upon reduction, these positive peaks disappear, and 2 negative bands appear per peak, corresponding to the reduced segments that previously made up the disulfide bonded peptides. Combining this result with the formation of 3 disulfide bonds upon folding, one can conclude that a third disulfide bridge is formed between Cys46 and Cys62. Partial tryptic digestion of the folded protein in 5M urea for 1 hour produced a fragment that contained Cys46 and Cys62 (amino acid residues 45-48 of SEQ ID NO: 4 and amino acid residues 49-64 of SEQ ID NO: 4). Reduction yielded a peptide fragment corresponding to amino acid residues 49-64. Longer digestion times in trypsin resulted in disulfide scrambling. Fig. 7 shows the disulfide bond map of the totally synthetic GLP-1R NTD.

Example 6
Fluorescence Anisotropy Binding Assay

Fluorescence anisotropy binding assays were performed as follows. 300 µl of a 0.7 µM solution of GLP-1 (7-36) labeled with tetramethyl rhodamine at Lys33 in binding buffer (125 mM Tris, pH 7.3, 150 mM NaCl, 1 mM EDTA) were placed into the thermostated (T = 25°C) sample compartment of a Fluorolog 3 L-format spectrofluorimeter with single excitation and emission spectrographs (ISA-Spex-Jobin-Yvon, New Jersey). For additional rejection of stray light, a 550 nm long-pass filter was added to the emission beam path. A stock solution of 300 µg of folded GLP-1R NTD in 50 µl binding buffer was prepared and added in small aliquots to the ligand solution. To account for non-specific binding, a stock solution of SDF-1alpha (a highly disulfide crosslinked chemokine) was prepared and the concentration was adjusted to be equivalent to the total molar concentration of amino acid residues.

Concentrations were determined by absorption spectroscopy. After each addition, the sample was allowed to equilibrate under stirring for 10 minutes at 25°C. Longer equilibration times did not lead to any significant changes in anisotropy. Fluorescence anisotropy and a magic-angle total emission scan from 590 nm to 630 nm were taken after excitation at 520 nm. Scans were taken with 4 nm step size and 1 s dwell time. To improve the S/N ratio, the total anisotropy from 590 nm to 630 nm was integrated for analysis.

Example 7
Binding Assay

Figs. 8A and 8B present the result of the anisotropy ligand-binding assay of the GLP-1 receptor in a semi-logarithmic representation. Clearly, beginning saturation of binding is observed with an approximate Kd50 of 17 μM. More interestingly, the sigmoidal shape of the ligand-binding curve in the linear representation suggests that binding of the receptor domain by the hormone is cooperative. Further studies to determine the exact stoichiometry of the ligand receptor complex, the extent of cooperativity and the binding constant are in progress.

The above Examples illustrate that chemical synthesis of membrane protein receptor domains can be utilized to provide facile and unprecedented access to the extramembranous domains of membrane protein receptors. The present invention opens the way for detailed structure-function studies of soluble receptor domains and for the development of homogenous and true high-throughput drug screening assays and diagnostics.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.